Bilinear Mixed Effects Models For Relations Between Universities

S. Alimoradi and M. Khalilian

Januarey 2008

Abstract

this article illustrates the use of linear and bilinear random effects models to represent statistical dependencies that often characterize dyadic data such as international relations. In particular, we show how to estimate models for dyadic data that simultaneously take into account: regressor variables and third-order dependencies, such as transitivity, clustering, and balance. We apply this new approach to the relations among ph.d. of university in Iran over the period from 1991-2005, illustrating the presence and strength of second and third-order statistical dependencies in these data.

1 Introduction

Social network data typically consist of a set of n actors and a relational tie $y_{i,j}$, measured on each ordered pair of actors $i, j = 1, \ldots, n$. This framework has many applications in the social and behavioral sciences including, for example, the behavior of epidemics, the interconnectedness of the World Wide Web, and telephone calling patterns.

In the simplest cases, $y_{i,j}$ is a dichotomous variable indicating the presence or absence of some relation of interest, such as friendship, collaboration, transmission of information or disease, and so forth. The data are often represented by an $n \times n$ sociomatrix Y. In the case of binary relations, the data can also be thought of as a graph in which the nodes are actors and the edge set is $\{(i, j) : y_{i,j} = 1\}$.Social network analysis is a broad area of social science research that has been developed to describe the relationships among interdependent units (Holland and Leinhardt 1971, Bondy and Murty 1976). It is somewhat surprising that to date there are no published applications using a social network framework to study international relations since it is evident at first blush that international politics is about the interdependencies that appear around the world. This is perhaps due to the fact that most tools for social network analysis are focused on the simple case of binary (0-1) relations, where the data can be represented by a simple graph (see Wasserman and Faust 1994, Wasserman and Pattison 1996). Dealing with non-binary data (such as counts or continuous data) or regressor variables has not been well addressed in the social networks literature (see Hoff, Raftery, and Handcock 2002 for a discussion). Herein, we develop a generalized regression framework for analyzing and accounting for the dependencies in valued and binary dyadic international relations data. This approach builds on the social relations model (Warner, Kenny and Stoto 1979; Wong 1982) that specifies random effects for the originator and recipient of a relation or action, as well as allowing for within dyad correlation of relations. We expand upon previous approaches by allowing for certain kinds of third-order dependence using an inner product of latent, unobserved characteristic vectors. The use of inner products to model dependencies is new, and related to the the recent development of "latent space" models for dyadic data (Hoff, Raftery and Handcock 2002).

2 Latent Space Approaches to Social Network Analysis

In some social network data, the probability of a relational tie between two individuals may increase as the characteristics of the individuals become more similar. A subset of individuals in the population with a large number of social ties between them may be indicative of a group of individuals who have nearby positions in this space of characteristics, or "social space". Various concepts of social space have been discussed by McFarland and Brown (1973) and Faust (1988). In the context of this article, social space refers to a space of unobserved latent characteristics that represent potential transitive tendencies in network relations. A probability measure over these unobserved characteristics induces a model in which the presence of a tie between two individuals is dependent on the presence of other ties. Relations modeled as such are probabilistically transitive in nature. The observation of $i \to j$ and $j \to k$ suggests that i and k are not too far apart in social space, and therefore are more likely to have a tie (Holland and Leinhardt 1971). In latent variable model it is assumed each actor i has an unknown position $\mathbf{z}_i$ in social space. The ties in the network are assumed to be conditionally independent given these positions, and the probability of a specific tie between two individuals is modeled as some function of their positions, such as the distance between the two actors in social space. Estimation of positions is simplified by the use of a logistic regression model, and confidence regions for latent positions are computable using standard MCMC algorithms.

2.1 Distance Models

We take a conditional independence approach to modeling by assuming that the presence or absence of a tie between two individuals is independent of all other ties in the system, given the unobserved positions in social space of the two individuals,

$$p(\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \underline{\theta}) = \prod_{i \neq j} p(y_{i,j}|z_i, z_j, x_{i,j}, \alpha, \beta)$$

where $\mathbf{X}$ and $x_{i,j}$ are observed characteristics which are potentially pair-specific and vector-valued and α, β and $\mathbf{Z}$ are parameters and positions to be estimated. A convenient parameterization is the logistic regression model in which the probability of a tie depends on the Euclidean distance between $\mathbf{z}_i$ and $\mathbf{z}_j$, as well as on observed covariates that $x_{i,j}$ measure characteristics of the dyad,

$$\eta_{i,j} = \log \frac{p(y_{i,j} = 1|z_i, z_j, x_{i,j}, \alpha, \beta)}{p(y_{i,j} = 0|z_i, z_j, x_{i,j}, \alpha, \beta)}$$

$$= \alpha + \beta' x_{i,j} - |z_i - z_j| \tag{1}$$

3 Linear Mixed Effects Models for Exchangeable Dyadic Data

Suppose we are only interested in estimating the linear relationships between responses $y_{i,j}$ and a possibly vector valued set of variables $x_{i,j}$, which could include characteristics of unit i, characteristics of unit j, or characteristics specific to the pair. In this case we might consider the regression model

$$y_{i,j} = \underline{\beta}' \mathbf{x}_{i,j} + \varepsilon_{i,j} \tag{2}$$

where $y_{i,i}$ is typically not defined. It is often assumed in regression problems that the regressors $x_{i,j}$ contain enough information so that the distribution of the errors is invariant under permutations

of the unit labels. This assumption is equivalent to the $n \times n$ matrix of errors (with an undefined diagonal) having a distribution that is invariant under identical row and column permutations, so that $\{\varepsilon_{i,j} \ : \ i \neq j\}$ is equal in distribution to $\{\varepsilon_{\pi(i),\pi(j)} \ : \ i \neq j\}$ for any permutation π of $\{1, \cdots, n\}$. This condition is called weak row-and- column exchangeability of an array. For undirected data, such exchangeability implies a "random effects" representation of the errors, in that

$$\varepsilon_{i,j} \sim f(\mu, a_i, a_j, \gamma_{i,j}) \tag{3}$$

where $\mu, a_i, a_j, \gamma_{i,j}$ are independent random variables and f is a function to be specified (Aldous 1985, Theorem 14.11). If in addition to the above invariance assumption we also model the errors as Gaussian, then the joint distribution can be represented in terms of a linear random effects model. In the more general case of directed observations, we can represent the joint distribution of the errors as follows:

$$\varepsilon_{i,j} = a_i + a_j + \gamma_{i,j} \tag{4}$$

where

$$a_1, \ldots, a_n \overset{i.i.d}{\sim} N(0 \ , \sigma^2_{\mathbf{a}})$$

$$(\gamma_{i,j}, \gamma_{j,i})' \sim MVN(\mathbf{0} \ , \mathbf{\Sigma}_{\underline{\gamma}}) \qquad , \qquad \mathbf{\Sigma}_{\underline{\gamma}} = \begin{pmatrix} \sigma^2_{\underline{\gamma}} & \sigma^2_{\underline{\gamma}} \\ \sigma^2_{\underline{\gamma}} & \sigma^2_{\underline{\gamma}} \end{pmatrix}$$

with effects otherwise being independent. The covariance structure of the errors (and thus the observations) is as follows:

$$E(\varepsilon^2_{i,j}) = \sigma^2_{\mathbf{a}} + \sigma^2_{\mathbf{b}} + \sigma^2_{\gamma} \quad , \quad E(\varepsilon_{i,j}\varepsilon_{i,k}) = \sigma^2_{\mathbf{a}}$$

$$E(\varepsilon_{i,j}\varepsilon_{j,i}) = \rho \ \sigma^2_{\gamma} + 2\sigma_{\mathbf{ab}} \quad , \quad E(\varepsilon_{i,j}\varepsilon_{k,j}) = \sigma^2_{\mathbf{b}}$$

$$E(\varepsilon_{i,j}\varepsilon_{k,l}) = 0 \quad , \quad E(\varepsilon_{i,j}\varepsilon_{k,i}) = \sigma_{\mathbf{ab}}$$

To analyze responses in particular sample spaces, the error structure described above can be added to a linear predictor in a generalized linear model:

$$\theta_{i,j} = \underline{\beta}'\mathbf{x}_{i,j} + a_i + b_j + \gamma_{i,j} \quad , \quad E(y_{i,j}|\theta_{i,j}) = g(\theta_{i,j})$$

This is a generalized linear mixed-effects model with inverse-link function $g(\underline{\theta})$, in which the observations are modeled as conditionally independent given the random effects, but are unconditionally dependent.

3.1 Modeling Third Order Dependence Patterns

Some dependence patterns commonly seen in dydaic datasets have been given the descriptive titles of balance and clusterability. for example after fitting a regression model and obtaining the residuals $\{\hat{\xi}_{i,j} \ : \ i \neq j\}$, the theoretic definitions of these concepts are as follows:

Definition 3.1 *For signed residuals, a triad i, j, k is said to be balanced if $\hat{\xi}_{i,j} \times \hat{\xi}_{j,k} \times \hat{\xi}_{i,k} > 0$*

Definition 3.2 *Clusterability is a relaxation of the concept of balance. A triad is clusterable if it is balanced or the relations are all negative. The idea is that a clusterable triad can be divided into groups where the measurements are positive within groups and negative between groups.*

Clusterability and balanced cycle of residuals are shown garaphically in Figure 1.

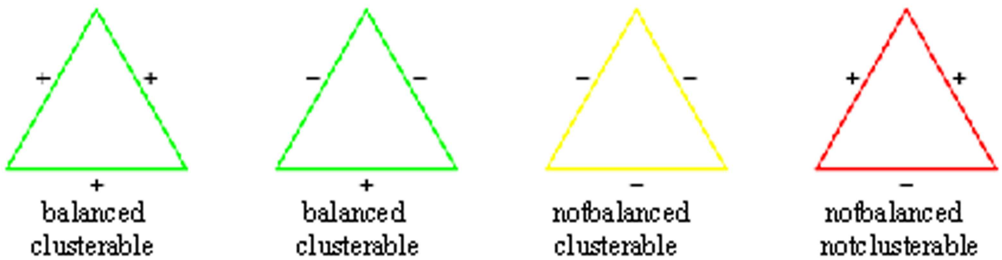


Figure 1 : Balance and Clusterability of Cycles

Hoff et al. (2002) used simple functions of latent characteristic vectors in a fixed effects setting to capture some forms of t balance and clusterability. For example, they considered models in which $\theta_{i,j} = \beta' x_{i,j} + f(\mathbf{z}_i, \mathbf{z}_j)$ where $f(\mathbf{z}_i, \mathbf{z}_j) = |\mathbf{z}_i - \mathbf{z}_j|$. we consider a similar approach using the inner product kernel $f(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{z}_i'\mathbf{z}_j$ and give random and fixed effects interpretations. Adding the bilinear effect to the linear random effects in models (4) gives

$$\varepsilon_{i,j} = a_i + a_j + \gamma_{i,j} + \xi_{i,j} \quad , \quad \xi_{i,j} = \mathbf{z}_i'\mathbf{z}_j \tag{5}$$

to suggest

$$\mathbf{z}_1, \ldots, \mathbf{z}_n \overset{i.i.d}{\sim} MVN(\mathbf{0}\ , \sigma^2_{\mathbf{z}}\mathbf{I}_{K\times K})$$

the nonzero second and third order moments are

$$E(\varepsilon^2_{i,j}) = 2\sigma^2_{\mathbf{a}} + \sigma^2_{\gamma} + k\sigma^4_{\mathbf{z}} \quad , \quad E(\varepsilon_{i,j}\varepsilon_{i,k}) = \sigma^2_{\mathbf{a}}$$

$$E(\varepsilon_{i,j}\varepsilon_{j,i}) = \sigma^2_{\gamma} + 2\sigma^2_{\mathbf{a}} + k\sigma^4_{\mathbf{z}} \quad , \quad E(\varepsilon_{i,j}\varepsilon_{k,j}) = \sigma^2_{\mathbf{a}}$$

$$E(\varepsilon_{i,j}\varepsilon_{j,k}\varepsilon_{k,i}) = k\sigma^6_{\mathbf{z}} \quad , \quad E(\varepsilon_{i,j}\varepsilon_{k,i}) = \sigma^2_{\mathbf{a}}$$

Thus the effect $\xi_{i,j} = \mathbf{z}_i'\mathbf{z}_j$ can be interpreted as a mean-zero random effect able to induce a particular form of third-order dependence often found in dyadic datasets.

4 Bilinear Mixed Effects Models Parameters Estimation

To obtain a "cleaner" partition of the variance and a more efficient MCMC sampling scheme,in model (2) we decompose $\mathbf{x}_{i,j}$ into $\mathbf{x}_{i,j} = (x_{i,j}, x_{\mathbf{s},i}, x_{\mathbf{s},j})$ i.e. into dyad specific regressors $x_{i,j}$, sender specific regressors $x_{\mathbf{s},i}$ and receiver specific regressors $x_{\mathbf{s},j}$. The generalized bilinear model is then rewritten as

$$\theta_{i,j} = \beta_d x_{i,j} + \underline{\beta}'_{\mathbf{s}}\mathbf{x}_{\mathbf{s},i} + \underline{\beta}'_{\mathbf{s}}\mathbf{x}_{\mathbf{s},j} + \varepsilon_{i,j}$$

or equivalently

$$\theta_{i,j} = \beta_d x_{i,j} + s_i + s_j + \gamma_{i,j} + \mathbf{z}_i'\mathbf{z}_j \tag{6}$$

$$s_i = \underline{\beta}'_{\mathbf{s}}\mathbf{x}_{\mathbf{s},i} + a_i$$

where $\mathbf{x}_{\mathbf{s},i} = (0/5, x_i)'$ and $\underline{\beta}_{\mathbf{s}} = (\beta_0, \beta_{\mathbf{s}})'$. This parameterization for the linear unit-level effects is similar to the "centered" parameterizations suggested by Gelfand et al. (1995, 1996). Note that an intercept can be thought of as both a sender or receiver specific effect. For symmetry, we include the constant 1/2 at the beginning of each $x_{\mathbf{s},i}$ and $x_{\mathbf{s},j}$ vector.

Using the above reparameterization for $\theta_{i,j}$, we estimate the parameters for the generalized bilinear

regression model by constructing a Markov chain in $\{\beta_d, \underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\mathbf{z}}, \mathbf{\Sigma}_{\underline{\gamma}}, \mathbf{Z}\}$ (where $\mathbf{Z}$ denotes the $k \times n$ matrix of latent vectors),having $p(\beta_d, \underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\mathbf{z}}, \mathbf{\Sigma}_{\underline{\gamma}}, \mathbf{Z})$ as the invariant distribution. This is obtained via an algorithm based on Gibbs sampling, which also samples **s**,**r** and the $\underline{\theta}$'s. The basic algorithm is to iterate the following steps:

1. Sample linear effeects:
   (a) Sample $\beta_d, \mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\mathbf{z}}, \mathbf{\Sigma}_{\underline{\gamma}}, \underline{\theta}, \mathbf{Z}$(linear regression);
   (b) Sample $\underline{\beta}_{\mathbf{s}}|\mathbf{s}, \sigma^2_{\mathbf{a}}$ (linear regression);
   (c) Sample $\sigma^2_{\mathbf{a}}$ and $\mathbf{\Sigma}_{\underline{\gamma}}$ from their full conditionals.

2. Sample bilinear effects:
   (a) For $i = 1, \ldots, n$ sample $\mathbf{z}_i|\{\mathbf{z}_j, j \neq i\}, \underline{\theta}, \underline{\beta}, \mathbf{s}, \sigma^2_{\mathbf{z}}, \mathbf{\Sigma}_{\underline{\gamma}}$ (a linear regression);
   (b) Sample $\sigma^2_{\mathbf{z}}$ from its full conditional.

3. Sample dyad specific parameters: Update $(\theta_{i,j}, \theta_{j,i})$ using a Metropolis-Hastings step:
   (a) Propose
   $$\begin{pmatrix} \theta^*_{i,j} \\ \theta^*_{j,i} \end{pmatrix} \sim \mathbf{MVN}\left(\begin{pmatrix} \underline{\beta}'\mathbf{x}_{i,j} + a_i + a_j + \mathbf{z}'_i\mathbf{z}_j \\ \underline{\beta}'\mathbf{x}_{j,i} + a_j + a_i + \mathbf{z}'_j\mathbf{z}_i \end{pmatrix}, \mathbf{\Sigma}_{\underline{\gamma}} \right)$$
   (b)Accept $\begin{pmatrix} \theta^*_{i,j} \\ \theta^*_{j,i} \end{pmatrix}$with probability
   $$\alpha = min\Big(\frac{P(y_{i,j}|\theta^*_{i,j})P(y_{j,i}|\theta^*_{j,i})}{P(y_{i,j}|\theta_{i,j})P(y_{j,i}|\theta_{j,i})}, 1\Big)$$

for more detail see Metropolis et al.(1953) and Hastings et al. (1970). Various combinations of the above steps can be used to estimate different models. The steps in 1 alone provide a Bayesian estimation procedure for the linear regression problem having an error covariance as in (2). Bayesian estimation of the normal bilinear model with the identity link could proceed by replacing each $\theta_{i,j}$ with $y_{i,j}$ and only iterating steps 1 and 2. Estimation of a generalized linear mixed effects model with random effects structure given by (2) could proceed by iterating steps 1 and 3. The full conditional distributions required to perform steps 1 and 2 are given below.

4.1 Conditional Distributions for the Linear Effects Components

Similar to Wong's (1982) approach to the invariant normal model, we let

$$u_{i,j} = \theta_{i,j} + \theta_{j,i} - 2\mathbf{z}'_i\mathbf{z}_j$$

$$v_{i,j} = \theta_{i,j} - \theta_{j,i} \quad for\ i < j$$

We then have

$$u_{i,j} = \beta_d(x_{i,j} + x_{j,i}) + 2(s_i + s_j) + \delta_{u_{i,j}} \quad , \quad \delta_{u_{i,j}} = \gamma_{i,j} + \gamma_{j,i}$$
$$v_{i,j} = 0$$

with definition $\mathbf{u} = \{u_{i,j}\}$, $\underline{\delta}_{\mathbf{u}} = \{\delta_{u_{i,j}}\}$ and $\mathbf{X}_{\mathbf{u}}$ the appropriate design matrices:

$$\mathbf{u} = \mathbf{X}_{\mathbf{u}} \begin{pmatrix} \beta_d \\ \mathbf{s} \end{pmatrix} + \underline{\delta}_{\mathbf{u}} \tag{7}$$

and

$$\mathbf{u} \sim MVN(\mathbf{X}_{\mathbf{u}}\mathbf{\Phi}, \sigma^2_{\mathbf{u}}\ \mathbf{I}_M) \quad , \quad \sigma^2_{\mathbf{u}} = 4\sigma^2_{\underline{\gamma}} \tag{8}$$

where $M = \frac{n(n-1)}{2}$ and $\mathbf{\Phi} = (\beta_d \ \mathbf{s}')'$. we have

$$\mathbf{s} \sim MVN(\mathbf{X_s}\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}\mathbf{I}_{n\times n}) \tag{9}$$

and $\mathbf{X_s} = (\mathbf{x}_{\mathbf{s},1}, \mathbf{x}_{\mathbf{s},2}, \ldots, \mathbf{x}_{\mathbf{s},n})'$. The full conditional distribution of model (6) is then

$$L(\mathbf{u}|\beta_d, \mathbf{s}, \sigma^2_{\underline{\gamma}}) \times L(\mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}})$$

$$\propto exp\left\{ -\frac{1}{2}\Big[(\ \mathbf{u} - \mathbf{X_u}\mathbf{\Phi}\ )'(\ \mathbf{u} - \mathbf{X_u}\mathbf{\Phi}\ )/\sigma^2_{\mathbf{u}} \right\}$$

$$\times\ exp\left\{ -\frac{1}{2}\Big[(\mathbf{s}'\mathbf{s} + \underline{\beta}'_{\mathbf{s}}\mathbf{X}'_{\mathbf{s}}\mathbf{X_s}\underline{\beta}_{\mathbf{s}} - 2\underline{\beta}'_{\mathbf{s}}\mathbf{X}'_{\mathbf{s}}\mathbf{s})/\sigma^2_{\mathbf{a}}\Big] \right\} \tag{10}$$

joint posterior distributions using approach Bayesian is then proportional to product of prior density and function likelihood (gelman 2003):

$$\pi(\mathbf{s}, \beta_d, \underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\underline{\gamma}}|\mathbf{u}) \propto$$

$$L(\mathbf{u}|\beta_d, \mathbf{s}, \sigma^2_{\underline{\gamma}})L(\mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}})\pi(\beta_d)\pi(\underline{\beta}_{\mathbf{s}})\pi(\sigma^2_{\mathbf{a}})\pi(\sigma^2_{\underline{\gamma})} \tag{11}$$

note that we assume the parameters is independent.

•Full conditional of $(\beta_d, \mathbf{s})$

The full conditional distribution of $(\beta_d, \mathbf{s})$ is then proportional to joint posterior density and obtain with omitting the terms that uncondition to $(\beta_d, \mathbf{s})$.

$$\pi(\beta_d, \mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\underline{\gamma}}, \mathbf{u}) \propto L(\mathbf{u}|\beta_d, \mathbf{s}, \sigma^2_{\underline{\gamma}})L(\mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}})\pi(\beta_d)$$

For a multivariate normal $(\mu_{\beta_d}, \sigma^2_{\beta_d})$ prior distribution on β_d and then with omitting the terms that uncondition to $(\beta_d, \mathbf{s})$:

$$\pi(\beta_d, \mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\underline{\gamma}}, \mathbf{u}) \propto$$

$$exp\left\{ \mathbf{\Phi}'\left[\begin{pmatrix} \mu_{\beta_d}/\sigma^2_{\beta_d} \\ \mathbf{X_s}\underline{\beta}_{\mathbf{s}}/\sigma^2_{\mathbf{a}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{u}/\sigma^2_{\mathbf{u}}\right] \right.$$
$$\left. -\frac{1}{2}\mathbf{\Phi}'\left[\begin{pmatrix} \sigma^{-2}_{\beta_d} & \mathbf{0} \\ \mathbf{0} & \sigma^{-2}_{\mathbf{a}}\mathbf{I}_{n\times n} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{X_u}/\sigma^2_{\mathbf{u}}\right]\mathbf{\Phi} \right\}$$

The conditional distribution is thus

$$\beta_d, \mathbf{s}|\underline{\beta}_{\mathbf{s}}, \sigma^2_{\mathbf{a}}, \sigma^2_{\underline{\gamma}}, \mathbf{u} \sim MVN(\underline{\mu}, \mathbf{\Sigma})$$

where

$$\underline{\mu} = \mathbf{\Sigma}\ \left[\begin{pmatrix} \mu_{\beta_d}/\sigma^2_{\beta_d} \\ \mathbf{X_s}\underline{\beta}_{\mathbf{s}}/\sigma^2_{\mathbf{a}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{u}/\sigma^2_{\mathbf{u}}\right]$$

and

$$\mathbf{\Sigma} = \left[\begin{pmatrix} \sigma^{-2}_{\beta_d} & \mathbf{0} \\ \mathbf{0} & \sigma^{-2}_{\mathbf{a}}\mathbf{I}_{n\times n} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{X_u}/\sigma^2_{\mathbf{u}}\right]^{-1}$$

•Full conditional of $\underline{\beta}_{\mathbf{s}}$
The full conditional distribution of $\underline{\beta}_{\mathbf{s}}$ is then proportional to joint posterior density and obtain with omitting the terms that uncondition to $\underline{\beta}_{\mathbf{s}}$. using (11)

$$\pi(\underline{\beta}_{\mathbf{s}}|\mathbf{s},\sigma^2_{\mathbf{a}},\mathbf{u}) \propto \ L(\mathbf{s}|\underline{\beta}_{\mathbf{s}},\sigma^2_{\mathbf{a}})\pi(\underline{\beta}_{\mathbf{s}})$$

For a multivariate normal on $\underline{\beta}_{\mathbf{s}}$ as follows $\underline{\beta}_{\mathbf{s}} \sim MVN\left(\underline{\mu}_{\underline{\beta}_{\mathbf{s}}}, \mathbf{\Sigma}_{\underline{\beta}_{\mathbf{s}}}\right)$ and then with omitting the terms that uncondition to $\underline{\beta}_{\mathbf{s}}$.

$$\underline{\beta}_{\mathbf{s}}|\mathbf{s},\sigma^2_{\mathbf{a}},\mathbf{u} \sim MVN(\underline{\mu},\mathbf{\Sigma})$$

where

$$\underline{\mu} = \mathbf{\Sigma}\ \left[\mathbf{\Sigma}^{-1}_{\underline{\beta}_{\mathbf{s}}}\underline{\mu}_{\underline{\beta}_{\mathbf{s}}} + \mathbf{X}'_{\mathbf{s}}\mathbf{s}/\sigma^2_{\mathbf{a}}\right] \quad , \quad \mathbf{\Sigma} = \left(\mathbf{\Sigma}^{-1}_{\underline{\beta}_{\mathbf{sr}}} + \mathbf{X}'_{\mathbf{s}}\mathbf{X}_{\mathbf{s}}/\sigma^2_{\mathbf{a}}\right)^{-1}$$

•Full conditional of $\sigma^2_{\mathbf{a}}$
The full conditional distribution of $\sigma^2_{\mathbf{a}}$ is then proportional to joint posterior density and obtain with omitting the terms that uncondition to $\sigma^2_{\mathbf{a}}$.

$$\pi(\sigma^2_{\mathbf{a}}|\mathbf{a}) \propto \ L(\mathbf{a}|\sigma^2_{\mathbf{a}})\pi(\sigma^2_{\mathbf{a}})$$

note that $a_1,\ldots,a_n \overset{i.i.d}{\sim} N(0\ ,\sigma^2_{\mathbf{a}})$

$$L(\mathbf{a}|\sigma^2_{\mathbf{a}}) \propto \ |\sigma^2_{\mathbf{a}}|^{-\frac{n}{2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n} a_i^2/\sigma^2_{\mathbf{a}}\right\}$$

For a inverse gamma distribution on $\sigma^2_{\mathbf{a}}$ as follows $\sigma^2_{\mathbf{a}} \sim IG(\alpha_{\mathbf{a}1},\alpha_{\mathbf{a}2})$ The full conditional distribution of $\sigma^2_{\mathbf{a}}$ is then

$$\sigma^2_{\mathbf{a}}|\mathbf{a} \sim IG\Big(\alpha_{\mathbf{a}1} + \frac{1}{2}n\ ,\ \alpha_{\mathbf{a}2} + \sum_{i=1}^{n} a_i^2/\sigma^2_{\mathbf{a}}\Big)$$

•Full conditional of $\sigma^2_{\underline{\gamma}}$
note that

$$\sigma^2_{\underline{\gamma}} = \sigma^2_{\mathbf{u}}/4$$

to find The full conditional distribution of $\sigma^2_{\mathbf{u}}$ using (11)

$$\pi(\sigma^2_{\mathbf{u}}|\beta_d,\mathbf{s},\sigma^2_{\underline{\gamma}},\mathbf{u}) \propto \ L(\mathbf{u}|\beta_d,\mathbf{s},\sigma^2_{\underline{\gamma}})\pi(\sigma^2_{\mathbf{u}})$$

For a inverse gamma distribution on $\sigma^2_{\mathbf{u}}$ as follows $\sigma^2_{\mathbf{u}} \sim IG(\alpha_{\mathbf{u}1},\alpha_{\mathbf{u}2})$ and with omitting the terms that uncondition to $\sigma^2_{\mathbf{u}}$ .The full conditional distribution of $\sigma^2_{\mathbf{u}}$ is then

$$\sigma^2_{\mathbf{u}}|\beta_d,\mathbf{s},\sigma^2_{\underline{\gamma}},\mathbf{u} \sim IG\Big(\alpha_{\mathbf{u}1} + \frac{1}{2}M\ ,\ \alpha_{\mathbf{u}2} + (\ \mathbf{u} - \mathbf{X}_{\mathbf{u}}\mathbf{\Phi}\ )'(\ \mathbf{u} - \mathbf{X}_{\mathbf{u}}\mathbf{\Phi}\ )\Big)$$

4.2 Conditional distributions for the Bilinear Effects Component

Let $e_{i,j} = (\theta_{i,j} + \theta_{j,i} - \hat{u}_{i,j})/2$, the residual of the symmetric part of the matrix of $\underline{\theta}$'s after fitting the linear effects, and let $\delta_{\mathbf{u},i,j} = \gamma_{i,j} + \gamma_{j,i}$. Considering the full conditional of $\mathbf{z}_i$, we have

$$\begin{aligned} e_{i,1} &= \mathbf{z}'_i\mathbf{z}_1 + \delta_{\mathbf{u},i,1}/2 \\ e_{i,2} &= \mathbf{z}'_i\mathbf{z}_2 + \delta_{\mathbf{u},i,2}/2 \\ &\vdots \\ e_{i,n} &= \mathbf{z}'_i\mathbf{z}_n + \delta_{\mathbf{u},i,n}/2 \end{aligned} \tag{12}$$

can write the equations to face matrix:

$$\mathbf{e}_{i,-i} = \mathbf{Z}'_{-i}\mathbf{z}_i + \frac{1}{2}\delta_{i,-i} \tag{13}$$

where $\mathbf{e}_{i,-i}$ errors vector to face $\{e_{i,j} \ : \ i \neq j\}$ and $\mathbf{Z}_{-i}$ matrix $k \times (n-1)$ obtain to omit of i column $\mathbf{Z}$. for example for $i = 1$:

$$\mathbf{e}_{1,-1} = \begin{pmatrix} e_{1,2} \\ e_{1,3} \\ \vdots \\ e_{1,n} \end{pmatrix} \qquad , \qquad \mathbf{Z}_{-1} = \begin{pmatrix} \mathbf{z}'_2 \\ \mathbf{z}'_3 \\ \vdots \\ \mathbf{z}'_n \end{pmatrix}'$$

note that $Var(\delta_{\mathbf{u},i,j}/2) = \sigma^2_{\mathbf{u}}/4$ likelihood function model (13)is then:

$$L(\mathbf{e}_{i,-i}|\mathbf{Z}_{-i}, \mathbf{z}_i, \sigma^2_{\mathbf{u}}) \propto exp\left\{ -\frac{1}{2}\Big[4(\,\mathbf{e}_{i,-i} - \mathbf{Z}'_{-i}\mathbf{z}_i\,)'(\,\mathbf{e}_{i,-i} - \mathbf{Z}'_{-i}\mathbf{z}_i\,)/\sigma^2_{\mathbf{u}}\Big]\right\}$$

posterior distributions $\mathbf{z}_i$ is proportional to product of prior density and function likelihood. to assume $\mathbf{z}_i \sim MVN(\mathbf{0}, \mathbf{\Sigma_z})$

$$\pi(\mathbf{z}_i|\mathbf{Z}_{-i}, \sigma^2_{\mathbf{u}}, \mathbf{\Sigma_z}) \propto L(\mathbf{e}_{i,-i}|\mathbf{Z}_{-i}, \mathbf{z}_i, \sigma^2_{\mathbf{u}})\pi(\mathbf{z}_i)$$

•Full conditional of $\mathbf{z}_i$

The full conditional distribution of $\mathbf{z}_i$ is then proportional to joint posterior density and obtain with omitting the terms that uncondition to $\mathbf{z}_i$.

$$\pi(\mathbf{z}_i|\mathbf{Z}_{-i}, \sigma^2_{\mathbf{u}}, \mathbf{\Sigma_z}) \propto exp\left\{ -\frac{1}{2} \times 4\Big(\Big[\mathbf{z}'_i\mathbf{Z}_{-i}\mathbf{Z}'_{-i}\mathbf{z}_i/\sigma^2_{\mathbf{u}}\Big] - \Big[2\mathbf{z}'_i\mathbf{Z}_{-i}\mathbf{e}_{i,-i}/\sigma^2_{\mathbf{u}}\Big]\Big)\right\}$$

$$\times exp\left\{ -\frac{1}{2}\Big[(\,\mathbf{z}'_i\mathbf{\Sigma}^{-1}_{\mathbf{z}}\mathbf{z}_i\,)\Big]\right\}$$

for other hands

$$\pi(\mathbf{z}_i|\mathbf{Z}_{-i}, \sigma^2_{\mathbf{u}}, \mathbf{\Sigma_z}) \propto \ exp\left\{\mathbf{z}'_i\Big[4\mathbf{Z}_{-i}\mathbf{e}_{i,-i}/\sigma^2_{\mathbf{u}}\Big] - \frac{1}{2}\mathbf{z}'_i\Big[\mathbf{\Sigma}^{-1}_{\mathbf{z}} + 4\mathbf{Z}_{-i}\mathbf{Z}'_{-i}/\sigma^2_{\mathbf{u}}\Big]\mathbf{z}_i\right\}$$

the full conditional of $\mathbf{z}_i$ is multivariate normal $(\underline{\mu}, \mathbf{\Sigma})$ with

$$\underline{\mu} = 4\mathbf{\Sigma}\mathbf{Z}_{-i}\mathbf{e}_{i,-i}/\sigma^2_{\mathbf{u}} \qquad , \qquad \mathbf{\Sigma} = \left(\mathbf{\Sigma}^{-1}_{\mathbf{z}} + 4\mathbf{Z}_{-i}\mathbf{Z}'_{-i}/\sigma^2_{\mathbf{u}}\right)^{-1}$$

4.3 Conditional distributions for the matrix covariance $\mathbf{\Sigma_z}$

to assume $\mathbf{z}_i \sim MVN(\mathbf{0}, \mathbf{\Sigma_z})$

$$L(\mathbf{z}_1, \ldots, \mathbf{z}_n|\mathbf{\Sigma_z}) \propto \ |\mathbf{\Sigma_z}|^{-\frac{n}{2}} exp\left\{ -\frac{1}{2}\ tr\mathbf{\Sigma}^{-1}_{\mathbf{z}}\mathbf{Z}'\mathbf{Z}\right\}$$

posterior distributions for $\mathbf{\Sigma_z}$ is proportional to product of prior density and function likelihood.

$$\pi(\mathbf{\Sigma_z}|\mathbf{Z}) \propto L(\mathbf{z}_1, \ldots, \mathbf{z}_n|\mathbf{\Sigma_z})\pi(\mathbf{\Sigma_z})$$

•Full conditional of $\mathbf{\Sigma_z}$

The full conditional distribution of $\mathbf{\Sigma_z}$ is then proportional to joint posterior density and obtain with

omitting the terms that uncondition to $\mathbf{\Sigma_z}$. to assume prior distributions, inverse wishart for $\mathbf{\Sigma_z}$ as follows $\mathbf{\Sigma_z} \sim \mathbf{IW}\,(\,\mathbf{\Sigma_{z0}}, \nu\,)$ we have

$$\pi(\mathbf{\Sigma_z}|\mathbf{Z}) \propto \; |\mathbf{\Sigma_z}|^{-\frac{(\nu+n)}{2}} exp\left\{ -\frac{1}{2}\; tr\mathbf{\Sigma_z}^{-1}\Big[\mathbf{\Sigma_{z0}} + \mathbf{Z}'\mathbf{Z}\Big]\right\}$$

to note that property of inverse wishart,The full conditional distribution of $\mathbf{\Sigma_z}$ is

$$\mathbf{\Sigma_z}|\mathbf{Z} \sim \mathbf{IW}\Big(\mathbf{\Sigma_{z0}} + \mathbf{Z}'\mathbf{Z}\;,\;\nu+n\Big)$$

Alternatively, if we restrict $\mathbf{\Sigma_z}$ to be $\sigma^2_{\mathbf{z}}\mathbf{I}_{k\times k}$ and use an inverse gamma for $\sigma^2_{\mathbf{z}}$ as follows $\sigma^2_{\mathbf{z}} \sim IG(\alpha_0, \alpha_1)$ and with omitting the terms that uncondition to $\sigma^2_{\mathbf{z}}$

$$\pi(\mathbf{\Sigma_z}|\mathbf{Z}) \propto \; \sigma_{\mathbf{z}}^{2-(\alpha_0+nk/2+1)} exp\left\{ -[\alpha_1 + tr\mathbf{Z}'\mathbf{Z}/2]/\sigma^2_{\mathbf{z}}\right\}$$

then

$$\sigma^2_{\mathbf{z}}|\mathbf{Z} \sim IG(\alpha_0 + nk/2\;,\;\alpha_1 + tr\mathbf{Z}'\mathbf{Z}/2)$$

4.4 Selecting the Latent Dimension

One issue in model fitting is the selection of the dimension k of the latent variables $\mathbf{z}$. Selection of K could depend on the goal of the analysis. For example, if the goal is descriptive, i.e. the desired end result is a decomposition of the variance into interpretable components, then a choice of $K = 1, 2$ or 3 would allow for a simple graphical presentation of a multiplicative component of the variance.Alternatively, one could examine model fit as a function of K based on the log-likelihood. having obtaind posterior estimates $\hat{\Psi}^{(k)} = \{\hat{\underline{\beta}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{Z}}, \hat{\mathbf{\Sigma}}_{\underline{\gamma}}\}$ for a range of K, one can compare the value of $\log p(\mathbf{Y}|\hat{\Psi}^{(k)})$ to assess model fit versus complexity. Az a funtion of K, the Akaike information criterion(AIC) and Bayesian information criterion (BIC) are

$$AIC(k) = -2\log p(\mathbf{Y}|\hat{\Psi}^{(k)}) + c + [2n] \times k$$

and

$$BIC(k) = -2\log p(\mathbf{Y}|\hat{\Psi}^{(k)}) + c + \Big[n\log\binom{n}{2}\Big] \times k$$

where the suggestion is to prefer the model with a lowest value of the criterion. for hierarchical model, Spiegelhalter, Best, Carlin, and van der Linde (2002) suggested usisng the deviance information criterion (DIC),

$$DIC(k) = -2\log p(\mathbf{Y}|\hat{\Psi}^{(k)}) + 2 \times p_D^{(k)}$$

where the penalty $p_D^{(k)}$ on the model complexity is given by

$$p_D^{(k)} = -2 \times \Big\{E\Big[\log p(\mathbf{Y}|\hat{\Psi}^{(k)})|\mathbf{Y}\Big] - \log p(\mathbf{Y}|\hat{\Psi}^{(k)})\Big\}$$

this expection can be approximated by averaging over MCMC samples. the penalty term $p_D^{(k)}$ has been referred to as the "effective number of parameters" because it has this interpretation in normal linear model.

5 Data Analysis: Relations Between Universities

for fitting bilinear mixed effect model, we analyze data on relations between 30 university in Iran. We take our response $y_{i,j}$ to be the total number of "positive" actions reportedly initiated by university i with target j from 1991 to 2005. Positive actions here include articles in connection statistics sciences that they have been published in Iranian Statistical Conference book. $x_{i,j}$ is the geographic distance between university i and j and x_i is log population(number of master in university). The occurrence of a action between any two given countries in these data is rare, with 86of the nondiagonal entries of the sociomatrix being equal to 0. some descriptive ploys of the raw data are given in Figure 2. Panel (a) plots the response on a log scale versus the geographic distance in thousands of kilometer between university i and j. More precisely, this distance is the minimum distance between two university, which is 0 if i and j in one city. On average, the number of action decreases as geographic distance increases. Panel (b) plots $\log(1+\sum_{j:j\neq i} y_{i,j})$, versus log population , wich suggests a positive relationship between response and population. The quantities $\sum_{j:j\neq i} y_{i,j}$ is typically called the outdegree of unit i.

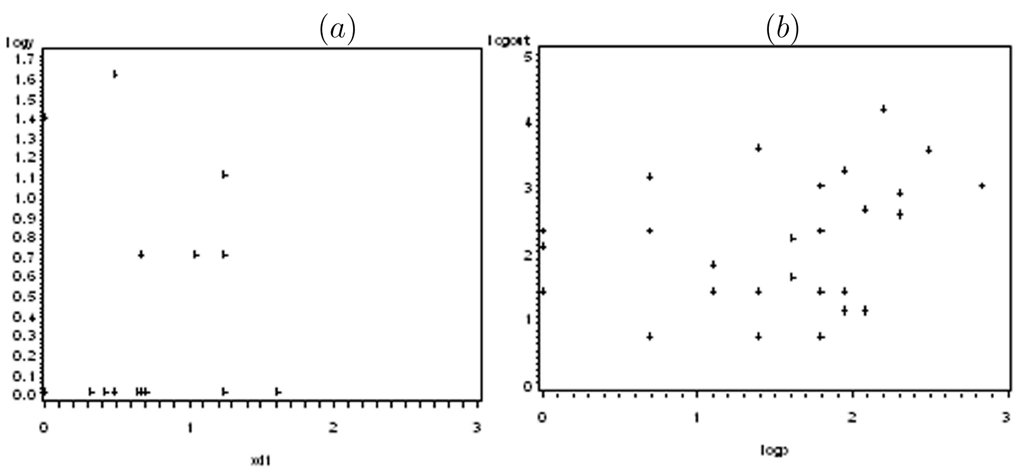


Figure 2 : Relationships Between (a)Response and Geographic Distance, and (b)Outdegree and Population

5.1 Evidence of Third-Order Dependence

Before fitting a somewhat complicated bilinear Poisson regression model, we evaluate the necessity of such an effort by looking for evidence of balance and clusterability in the data. we do this by fitting a simple linear regression on the logtransformed data and examining the residuals for third-order dependencies of the types described. more specifically, we obtain ordinary least squares estimates for the regression model

$$\log(y_{i,j}+1) = \beta_0 + \beta_d x_{i,j} + a_i + a_j + \xi_{i,j}$$

There are several indications of third-order dependence in these residuals:

1. Because the mean of the residuals is 0, independence of the residuals implies that the average value of the product $\hat{\xi}_{i,j}\hat{\xi}_{j,k}\hat{\xi}_{k,i}$ over triads also should be 0 (the concept of independence of the residuals is $E(\hat{\xi}_{i,j}\hat{\xi}_{j,k}\hat{\xi}_{k,i}) = 0$). As discussed in section 3.1, a value larger than 0 would indicate some degree of balance. The empirical average over triads turns out to be 0.0035.

2. The fraction of residuals that are positive is $p = 0.45$(the distribution of residuals is not symmetric). Under independence, the proportion of cycles that we would expect in the two balanced

categories shown in Figure 1 (+++ and +–) are $p^3 = 0.091$ and $3p(1-p)^2 = 0.4$, whereas the observed proportion are 0.115 and 0.385. The observed proportion in the unclusterable category (++-) is 0.333 and the value expected under independence is $3p^2(1-p) = 0.334$. The expected proportion in the clusterable but unbalance category is 0.166, and the observed proportion is 0.167.

3. As described in section 3.1, in a balance system we expect that if $\hat{\xi}_{i,j} > 0$, then $\hat{\xi}_{j,k}$ and $\hat{\xi}_{i,k}$ will have the same sign. such a pattern is shown graphically in Figure 3, which for each pair $\{i,j\}$ plots $\hat{\xi}_{i,j}$ versus the proportion of other nodes k for which $\hat{\xi}_{i,k} \times \hat{\xi}_{j,k} > 0$. Although the distribution of residuals is far from normal, we do see some indication of this type of third-order dependence. As we would expect from a balanced system, pairs $\{i,j\}$ for which $\hat{\xi}_{i,j}$ is less than 0 generally have dissimilar residuals to other,$\hat{P}(\hat{\xi}_{i,k} \times \hat{\xi}_{j,k} > 0)$ tends to be 0.47, pair $\{i,j\}$ for which $\hat{\xi}_{i,j}$ is greater than 0 generally have similar residuals to other, $\hat{P}(\hat{\xi}_{i,k} \times \hat{\xi}_{j,k} > 0)$tends to be greater than 0.52.

in the next section we analysis results of fitting bilinear mixed model.

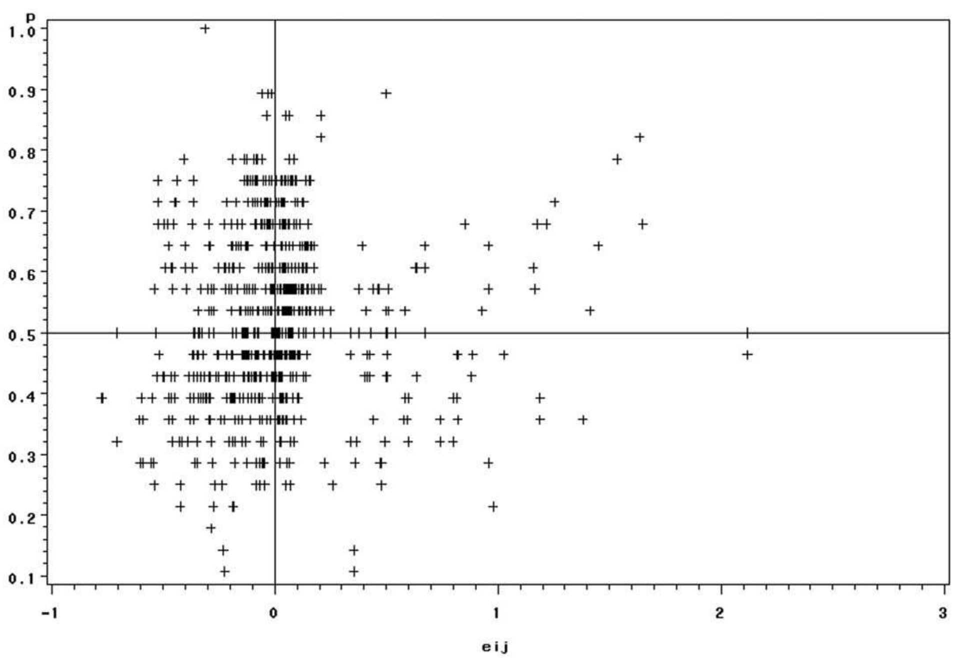


Figure 3 : *Balanced Residuals*

5.2 Model Selection

We fit the bilinear mixed effects model to the data using a Poisson distribution and the log-link, so that each response $y_{i,j}$ is assumed to have come from a Poisson distribution with mean $exp(\theta_{i,j})$, and that the $\mathbf{y}$'s are conditionally independent given the $\underline{\theta}$'s. assume following model:

$$\theta_{i,j} = \beta_d x_{i,j} + \underline{\beta}_{\mathbf{s}}' \mathbf{x}_{\mathbf{s},i} + \underline{\beta}_{\mathbf{s}}' \mathbf{x}_{\mathbf{s},j} + \varepsilon_{i,j}$$

where $\mathbf{x}_{\mathbf{s},i} = (0/5, x_i)'$ and $\underline{\beta}_s = (\beta_0, \beta_{\mathbf{s}})'$. Table 1 includes are the the marginal probability criteria, the DIC penalty,p_D, in the third column, the AIC criterion , the BIC criterion and the DIC criterion. In terms of the marginal likelihood criterion, the biggest improvements in fit are in going from $K = 1$ to $K = 2$ and from $K = 2$ to $K = 3$. Using the AIC criterion and penalizing the improvement in likelihood by the number of additional parameters, we would choose $K = 0$. The BIC, with a higher penalty on the number of parameters, favors $K = 0$. In contrast, the DIC favors $K = 1$.

Note that the increase in the DIC penalty tends to decrease the DIC. But note that for models whit $K = 1$ and $K = 2$ the DIC criterion have like predictive ability, then based on these results and our ability to plot in two dimensions, we choose to present the analysis of the $K = 2$ model in more detail.

K	$\log p(\mathbf{Y}\|\hat{\underline{\beta}}, \hat{\mathbf{s}}, \hat{\mathbf{Z}}, \hat{\sigma}^2_{\underline{\gamma}})$	P_D	$AIC(K)$	$BIC(K)$	$DIC(K)$
0	– 167.30	– 6.00	334.6	334.6	322.62
1	– 178.82	– 29.28	417.6	436.79	299.08
2	– 169.30	– 9.49	458.6	496.9	319.80
3	– 152.96	21.66	485.8	543.3	349.24
4	– 157.03	12.62	554.0	630.6	339.30

$Table\ 1 : Evaluation\ of\ K$

One Markov chains of length 100,000 was constructed using the algorithm described in section 4. The second chain used starting values obtained from the following procedure:

• fitting generalized linear model,using geographic distance as a regressor and sender and receiver labels as factor variables.

$$\log(y_{i,j} + 1) = \beta_d x_{i,j} + a_i + a_j + \xi_{i,j}$$

we let parameters of prior distribution for β_d to case $var(\hat{\beta}_d) = \sigma^2_{\beta_d}$ and $\mu_{\beta_d} = \hat{\beta}_d$

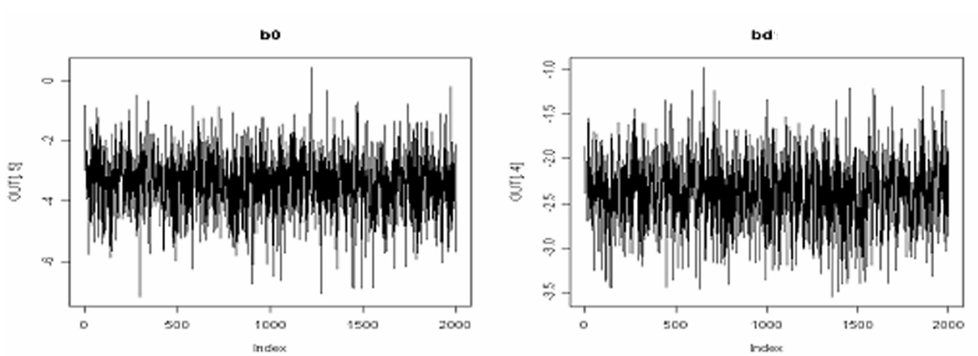
• letting $\hat{s}_i = (\hat{a}_i + \hat{a}_j)/2$ and fitting ordinary regression model we have

$$\hat{s}_i = \underline{\beta}'_{\mathbf{s}} \mathbf{x}_{\mathbf{s},i} + a_i$$

we let $\mathbf{\Sigma}_{\underline{\beta}_{\mathbf{s}}} = cov(\hat{\underline{\beta}}_{\mathbf{s}}), \underline{\mu}_{\underline{\beta}_{\mathbf{s}}} = \hat{\underline{\beta}}_{\mathbf{s}}$ and $(\alpha_{\mathbf{a}1}, \alpha_{\mathbf{a}2}) = (2, \hat{\sigma}^2_{\mathbf{a}})$

• The iteratively reweighted least-squares fitting procedure produces a matrix $\mathbf{R}$ of working residuals, with the of diagonal elements undefined. An estimate $\hat{\mathbf{Z}}$ of $\mathbf{Z}$ was then obtained by approximating $\mathbf{R}$ with a matrix product of the form $\mathbf{Z}'\mathbf{Z}$. This can be done with an iterative least-squares procedure, similar to the Gibbs sampling procedure outlined in Section 4: see ten Berge and Kiers (1989) for more details on this problem.

• An estimate of $\mathbf{\Sigma}_{\underline{\gamma}}$ is then obtained from $E = \mathbf{R} - \hat{\mathbf{Z}}'\hat{\mathbf{Z}}$. we define $\hat{\sigma}^2_{\mathbf{u}} = var(E + E')$ then $\hat{\sigma}^2_{\underline{\gamma}}$ update from $\hat{\sigma}^2_{\mathbf{u}}$.



$Figure\ 4 : Marginal\ MCMC\ output\ for\ parameters\ \beta_d$ and β_0

Samples of parameter values were saved from the Markov chains every 100 iterations, and for example for β_d and β_0 are plotted in Figures 4. The chain appear to have achieved stationarity after about 10,000 iterations, and so we base our inference on the saved samples after this point.
Posterior means and standard deviations of the model parameters, based on the saved MCMC samples are given in Table 2.

As in the raw data, we see a negative relation between response and geographic distance $E[\beta_d|\mathbf{Y}] = -2.38$, and a positive relation between response and log number of master in university ($E[\beta_s|\mathbf{Y}] = 0.65$).

	β_d	β_0	$\beta_{\mathbf{s}}$	$\sigma^2_{\mathbf{a}}$	$\sigma^2_{\underline{\gamma}}$	$\sigma^2_{\mathbf{z}_1}$	$\sigma^2_{\mathbf{z}_2}$
MEAN	– 2.38	– 3.41	0.65	1.1	0.86	0.64	0.52
SD	0.38	0.98	0.29	0.41	0.48	0.32	0.31

Table 2 : *Posterior means and standard deviations for* $k = 2$

Next, we analyze the posterior distribution of the the $k \times n$ matrix of latent vectors $\mathbf{Z}$. In Figure 5 has ploted sample $\mathbf{z}$'s over the plot of the means. Generally, two university will be modeled as having $\mathbf{z}$ in the same direction if they have large responses to one another relative to their total number of actions and covariate values, and/or if their responses involving other university are similar. For example, 3 university Shiraz, Olum pezeshky Shiraz and Azad Shiraz have placed in the same direction and so these university are similar because they had minimum 3 partner article, also Olum pezeshky Shiraz and Azad Shiraz did not have connection with other university exept Shiraz. Mashhad university had contacted at least with 12 university, thus it had reposed in central of plot. university Tarbiat Modares in addition to connections with other university, it had 11 partner article with Azad Oloum Tahghighat so this two university have placed in the same direction.

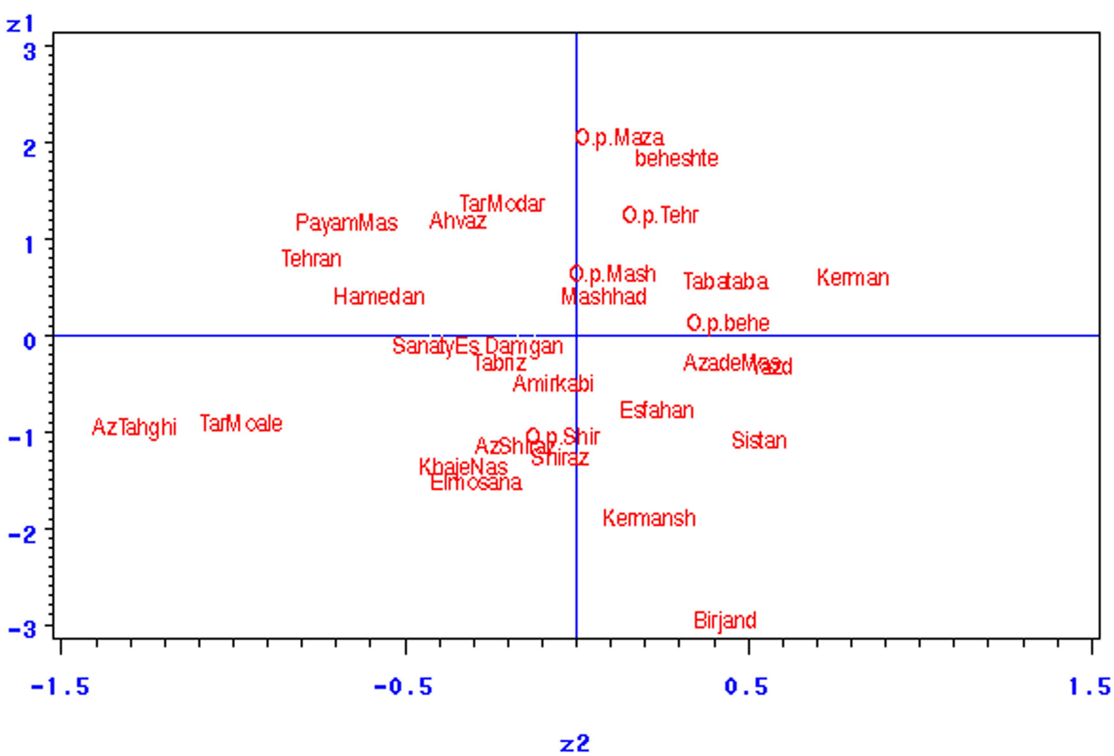


Figure 5 : *Posterior mean of* Z

References

[1] Aldous,D.J.(1985),Exchangeability and related topics, in Ecole dete de probabilites de Saint-Flour, XIII-1983, vol. 1117 of Lecture Notes in Math., pp. 1-198, Springer, Berlin.

[2] Bondy A., Murty U.S.R.(1976), Graph Theory with Applications, North Holland, New York.

[3] Faust K.(1988), Comparison of Methods for Positional Analysis: Structural and General Equivalence, Social Networks, **10**:313-341.

[4] Gelfand A.E., Sahu S.K., Carlin B.P. (1995), Efficient parameterisations for normal linear mixed models, Biometrika **82** 479-488.

[5] Gelfand A.E., Sahu S.K., Carlin B.P.(1996),Efficient parametrizations for generalized linear mixed models, in Bayesian statistics**5**, eds.J. Bernardo et al., New York: Oxford Univ. Press, pp 165-180.

[6] Gelman A., Carlin J.B., Stern H.S., Rubin D.B.(2003), Bayesian data analysis, 2nd Edition, Chapman and Hall, London.

[7] Hastings W. K.(1970), Monte Carlo sampling Methods using Markov Chains and their Applications, Biometrika **57** : 97-109.

[8] Hoff P.D., Raftery A.E., Handcock M.S.(2002), Latent space approaches to social network analysis, Journal of the American Statistical Association, **97** 1090-1098.

[9] Hoff P.D.(2005), Bilinear Mixed Effects Models for Dyadic Data, , Journal of the American Statistical Association, **100** N0.469: 286-295.

[10] Holland P.W., Leinhardt S.(1971). Transitivity in structural models of small groups. Comparative Group Studies **2**: 107-124.

[11] McFarland D., Brown D.(1973), Social Distance as Metric: A Systematic Introduction to Smallest Space Analysis, in Bonds of Pluralism: The Form and Substance of Urban Social Networks, ed. E. Laumann, New York: Wiley, pp. 213-253.

[12] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. (1953), Equations of state calculations by fast computing machines, Journal of Chemical Physics, **21**:1087-1091.

[13] Spiegelhalter D.J. , Best N.G , Carlin B.P, van der Linde A.(2002), Bayesian Measures of Model Complexity and Fit, Journal of the Royal Statistical Society. Series B (Statistical Methodology), **64**, No.4: 583-639.

[14] Warner R., Kenny D.A., Stoto M.(1979), A new round robin analysis of variance for social interaction data, Journal of Personality and Social Psychology 37:1742-1757.

[15] Wasserman S., Faust K.(1994), Social Network Analysis, Cambridge University Press.

[16] Wasserman S., Pattison P.(1996), Logit models and logistic regressions for social networks:I. An introduction to Markov random graphs and p^*, Psychometrika **61**: 401-426.

[17] Wong G.Y.(1982), Round Robin Analysis of Variance via Maximum Likelihood, Journal of the American Statistical Association **77**: 714-724.